

NASA CR-177826

Research Publication 106/OR
May 1981

E86-10020
NASA-CR-177826

DEVELOPMENT OF COMPUTER SOFTWARE TO ANALYZE
ENTIRE LANDSAT SCENES AND TO SUMMARIZE
CLASSIFICATION RESULTS OF
VARIABLE-SIZE POLYGONS

(E86-10020 NASA-CR-177826) DEVELOPMENT OF
COMPUTER SOFTWARE TO ANALYZE ENTIRE LANDSAT
SCENES AND TO SUMMARIZE CLASSIFICATION
RESULTS OF VARIABLE-SIZE POLYGONS Final
Report (Pennsylvania State Univ.) 29 p
Principal Investigator

N86-16688

Unclas
G3/43 00020

Brian J. Turner

Contributors

George M. Baumer

Wayne L. Myers

Steven G. Sykes

Office for Remote Sensing of Earth Resources



INSTITUTE FOR RESEARCH ON LAND & WATER RESOURCES

Original photography may be purchased
from EROS Data Center
Sioux Falls, SD 57198

THE PENNSYLVANIA STATE UNIVERSITY
UNIVERSITY PARK, PENNSYLVANIA

DEVELOPMENT OF COMPUTER SOFTWARE TO ANALYZE
ENTIRE LANDSAT SCENES AND TO SUMMARIZE
CLASSIFICATION RESULTS OF
VARIABLE-SIZE POLYGONS

Principal Investigator

Brian J. Turner

Contributors

George M. Baumer

Wayne L. Myers

Steven G. Sykes

The Pennsylvania State University, in compliance with federal and state laws and regulations governing affirmative action and nondiscrimination, does not discriminate in the recruitment, admission, and employment of students, faculty, and staff in the operation of any of its educational programs and activities as defined by law. Accordingly, nothing in this publication should be viewed as directly or indirectly expressing any limitation, specification, or discrimination as to race, religion, color, or national origin; or to handicap, age, sex, or status as a disabled or Vietnam-era veteran, except as provided by law. Inquiries concerning this policy may be directed to the Affirmative Action Officer.

Final Report: NASA Contract NAS5-26166

Prepared for

Goddard Space Flight Center
Greenbelt, Maryland 20771

TABLE OF CONTENTS

	Page
INTRODUCTION	1
SPECIFIC OBJECTIVES	3
RESULTS	4
Upgrade of Programs for Processing Full Landsat Scenes	4
Tabulation of Classification Results for Irregular-Shaped Polygons Within a Scene	5
General Method	6
Processing Digitizes Output	8
Arc Cleaning	8
Plotting	12
Interactive Arc File Editor	15
Polygon Formation	15
Map Registration	17
Compression and Decompression of Coordinates	17
Map File Subsets	18
Polygon-to-Raster Conversion	18
Color Display	19
Overlay of Ancillary Cartographic Data	19
Gray-Scale Display	20
Determination of Area Statistics	20
Feasibility of a Statewide Data Base/Information System	23
CONCLUSIONS	25
REFERENCES	26

INTRODUCTION

The Forest Pest Management Division (FPMD) of the Pennsylvania Bureau of Forestry has the responsibility for conducting annual surveys of the State's forest lands to accurately detect, map, and appraise forest insect infestations. This has proven to be an overwhelming and impractical task using current methods, which primarily consist of aerial sketch mapping and ground observations. Nevertheless, these surveys are vital to provide the spatial information needed to plan, organize, direct, and carry out control measures. These measures include the optimum use and allocation of pesticides and the introduction of natural predators.

A standardized, timely, and cost-effective method of accurately surveying forests and their condition should enhance the probability of suppressing infestations. The repetitive and synoptic coverage provided by Landsat (formerly ERTS) makes such satellite-derived data potentially attractive as a survey medium for monitoring forest insect damage over large areas. Forest Pest Management Division personnel have expressed keen interest in Landsat data and have informally cooperated with NASA/Goddard Space Flight Center (GSFC) since 1976 in the development of techniques to facilitate their use. The results of this work indicate that it may be feasible to use Landsat digital data to conduct annual surveys of insect defoliation of hardwood forests.

Since the first Landsat data for Pennsylvania became available in 1972, it has been apparent that large contiguous areas of heavy gypsy moth defoliation could be identified and mapped from satellite passes made in late June or early July. In fact, Mr. Darrel Williams, working with the Office for Remote Sensing of Earth Resources (ORSER) was granted a Master of Science degree at The Pennsylvania State University in 1974 with a thesis entitled, "Computer Analysis and Mapping of Gypsy Moth Defoliation Levels in Pennsylvania Using ERTS-1 Data." Further research by Mr. Williams at GSFC and by Dr. Brian Turner at ORSER indicated however, that if this technology was to reach an operational stage where the whole state could be scanned for accurate identification of defoliated areas, considerable gains in efficiency of computer processing and reductions in cost would be necessary.

A three-stage, three-year project was thus instituted by NASA, to see whether these gains could be achieved. The specific objectives of this project are to:

1. Demonstrate the feasibility of conducting automated, annual assessments of the acreage and severity of insect defoliation of hardwood forests using Landsat digital data.
2. Evaluate the accuracy, timeliness and cost-effectiveness of using the automated, Landsat-based survey approach, and compare these methods with current survey techniques.

3. Provide FPMD personnel with training and experience in the analysis of remote sensing data.
4. Assist ORSER in the development and implementation of computer software to facilitate:
 - a. ingestion and analysis of entire Landsat scenes, and
 - b. summarization of classification results for any given shape or size polygon (i. e., county or district boundaries) within a scene.

The project is being conducted in three phases:

- I. a preliminary testing, training, and development phase conducted within at least two county-wide study areas;
- II. a quasi-operational testing phase, operating within the framework of entire Landsat scenes; and
- III. a functional technology transfer phase.

Phase I, the preliminary test phase, was designed to demonstrate the practicality and accuracy of estimating the acreage and severity of defoliation using Landsat digital data and computer processing techniques. The primary objectives of Phase I were to:

1. Evaluate the accuracy, timeliness and cost effectiveness of assessing defoliation damage using Landsat data in comparison to the survey techniques currently used by the FPMD.
2. Define the Landsat-based analysis techniques to meet FPMD requirements.
3. Provide training to FPMD personnel relative to the digital analysis of remote sensing data.
4. Initiate the development and implementation of computer software on Penn State University computers in order to upgrade the ORSER digital image analysis package to allow both the processing of entire Landsat scenes and the tabulation of classification results for any irregular-shaped polygonal area within a scene.

SPECIFIC OBJECTIVES

A contract was developed with GSFC for ORSER to implement the fourth objective of Phase I. In this contract, ORSER was required to:

1. Initiate the development and implementation of computer software on The Pennsylvania State University computers in order to upgrade the ORSER digital image analysis package to allow both the processing of entire Landsat scenes and the tabulation of classification results for any irregular-shaped polygon within that scene.
2. Examine the feasibility of developing a data base/information system to incorporate Landsat and ancillary data covering the entire state of Pennsylvania.

RESULTS

Results of work completed under this contract have been realized in three major areas: expansion of the ORSER system to handle entire Landsat scenes, tabulation of classification results for irregular-shaped polygons within a scene, and assessment of the feasibility of a state-wide data base/information system incorporating Landsat data.

Upgrade of Programs for Processing Full Landsat Scenes

The ORSER format is basically a band-interleaved-by-line format, where all bands are stored on the same physical record from any line. The original limit of 3696 bytes per line permitted storage of fewer than 800 pixels for a five-band data set. This made working with entire scenes awkward, because they had to be divided into four or more vertical panels. The length of the data set was no problem, since it was limited only by the storage medium.

Two options for upgrading the ORSER system programs were considered. The first was to extend the buffer size of selected programs. This had the advantage that it would require less programming effort than the alternative, but it had the disadvantage of requiring more memory partitions, thus pushing jobs into lower priority categories and increasing turnaround time.

The second alternative was to alter the ORSER data format so that line lengths could be split into successive, or "continuation," records. This had the advantage that lines of any length could be accommodated without increasing buffer size but had the disadvantage that considerable rewriting of all programs would be necessary.

A meeting was held at Goddard Space Flight Center on May 30, 1980, at which time it was agreed that ORSER would follow the first option, and increase the buffer size of programs approximately fivefold, to accommodate larger data sets. The following programs (in order of priority) were to be extended:

- | | | |
|------------|--------------|---------------|
| (1) SUBSET | (6) CLASS | (11) SUBTRAN |
| (2) NMAP | (7) CLUS | (12) SUBBOUND |
| (3) SPAN | (8) RATIO | (13) SUBGM |
| (4) UMAP | (9) HISTRAN | |
| (5) USTATS | (10) DISPLAY | |

SUBGM was given lowest priority because GSFC has agreed to supply ORSER with geometrically corrected tapes for the contract activities.

The five-fold increase was implemented by increasing the buffer-size from 3696 bytes to 18,410 bytes per line. The limit of 18,410 bytes was chosen because it is large enough to handle a full Landsat scene of five bands yet small enough to allow most of the programs to run using a

relatively small amount of computer memory. The expanded programs now range in size from 300K to 800K.

In addition to those specified in the contract, the following programs were also expanded.

- | | |
|---------------|--------------|
| (14) REFORMAT | (16) MAPCOMP |
| (15) MERGE | (17) TPINFO |

The REFORMAT program was expanded because the raw data supplied by GSFC will be in a band-sequential format and REFORMAT is the program that reads this type of format and outputs in ORSER format. The MAPCOMP and MERGE programs were expanded so that the two different approaches to change-detection could be used--an important aspect of monitoring defoliation of forests. The TPINFO program was expanded for user convenience.

For a number of reasons, the expansion process is much more difficult than is apparent at first. The ORSER system was written by a number of different people and, although effort was made to make the programs consistent in their philosophy of operation, differences do exist and therefore each program had to be treated individually. This is no small task, since altogether the system consists of about 47,000 source statements.

Making a rather large modification to a rather large program usually introduces some unanticipated "bugs." Extensive testing has been done to try to debug the programs before non-ORSER users are exposed to the system, but obviously not every option of every program could be tested. This testing is an ongoing process and progressing well. It is expected that by June 30, 1981, the entire ORSER system will be expanded to handle the larger format.

Tabulation of Classification Results for Irregular-Shaped Polygons Within a Scene

In order to be able to tabulate classification results for irregular-shaped polygons within a scene, it was first necessary to develop a means of digitizing boundary lines from an existing map, form polygons, and then convert the polygons to raster form so that they could be overlaid on a raster map of classified data. The overlay process could be achieved with the existing MAPCOMP program, which would also produce area statistics for the intersection of the two maps.

Work had already begun on a rudimentary system for digitization when the current contract went into effect. Under the contract, the system was expanded and made more versatile and more user-friendly through the use of Penn State's conversational INTERACT editing system. For completeness, the whole system, which will form part of a master's thesis by Mr. Steven Sykes, is described here.

General Method

The method devised for digitizing maps uses the topological elements of a polygon map, i.e., arcs, nodes, and polygons (Fig. 1). This keeps the number of points that have to be digitized to a minimum and also reduces the space needed to store maps. The only points that are duplicated in a map file are nodes--the points where boundaries intersect. This method is similar to one that has been used at the U.S. Geological Survey (Fegeas, 1977).

Three types of arcs can be found on a typical thematic map (see Fig. 1). The most common arc defines a boundary between classes and separates two uniform mapping units throughout its entire length. A second type, the boundary arc, separates an individual mapping unit from the non-map area. These arcs make up the border of the digitized map. A third type, the connector arc, must be digitized when one mapping unit totally surrounds one or more other units. Connector arcs may be arbitrarily placed to link the surrounded mapping units, or islands, with the outside edge of the surrounding unit. Although connector arcs are necessary for area computations and topological verification of the digitized map, they need not appear in plots. All three arc types are readily identifiable within map files. Boundary arcs have one class on the right side and non-map on the left. Connector arcs have identical left and right values. The remainder of the arcs define different classes on the left and right hand sides. Plots may be easily made for any or all of these arc types.

Nodes are defined as the junction of at least three arc endpoints. They are the only points which are repeated both in digitizing and storage. If only two arcs terminate at a node, then the node is actually an internal point in what should be a single arc. Nodes are located anywhere boundary lines intersect on a map.

Polygons, the actual mapping units, are composed of one or more arcs. Polygons are of two general types, island and non-island, with the vast majority being non-island. Simple island polygons exist when a single mapping unit is surrounded by another single mapping unit. Complex islands exist when a grouping of contiguous mapping units are surrounded by a single mapping unit. Because arcs that define simple island polygons are loops, an arbitrary point is selected to be both the beginning and end when the arc is digitized. Island polygons are linked with the rest of the map by connector arcs.

The link between class attributes, or properties, and the maps can be established once the mapping unit on the left and right hand side of each arc is known. Digitizing a single point contained anywhere inside each mapping unit is all that is needed to have this link established. Files are created which contain the X-Y coordinates of these single points and their attribute, or mapping unit. Mapping unit attributes must be manually added to this file, but computer software can determine exactly which attributes belong to each digitized arc and append a numerical code to represent them. Because processing is done with half-word, or 16-bit, integer variables to represent mapping units, up to 65,536 different attributes are possible.

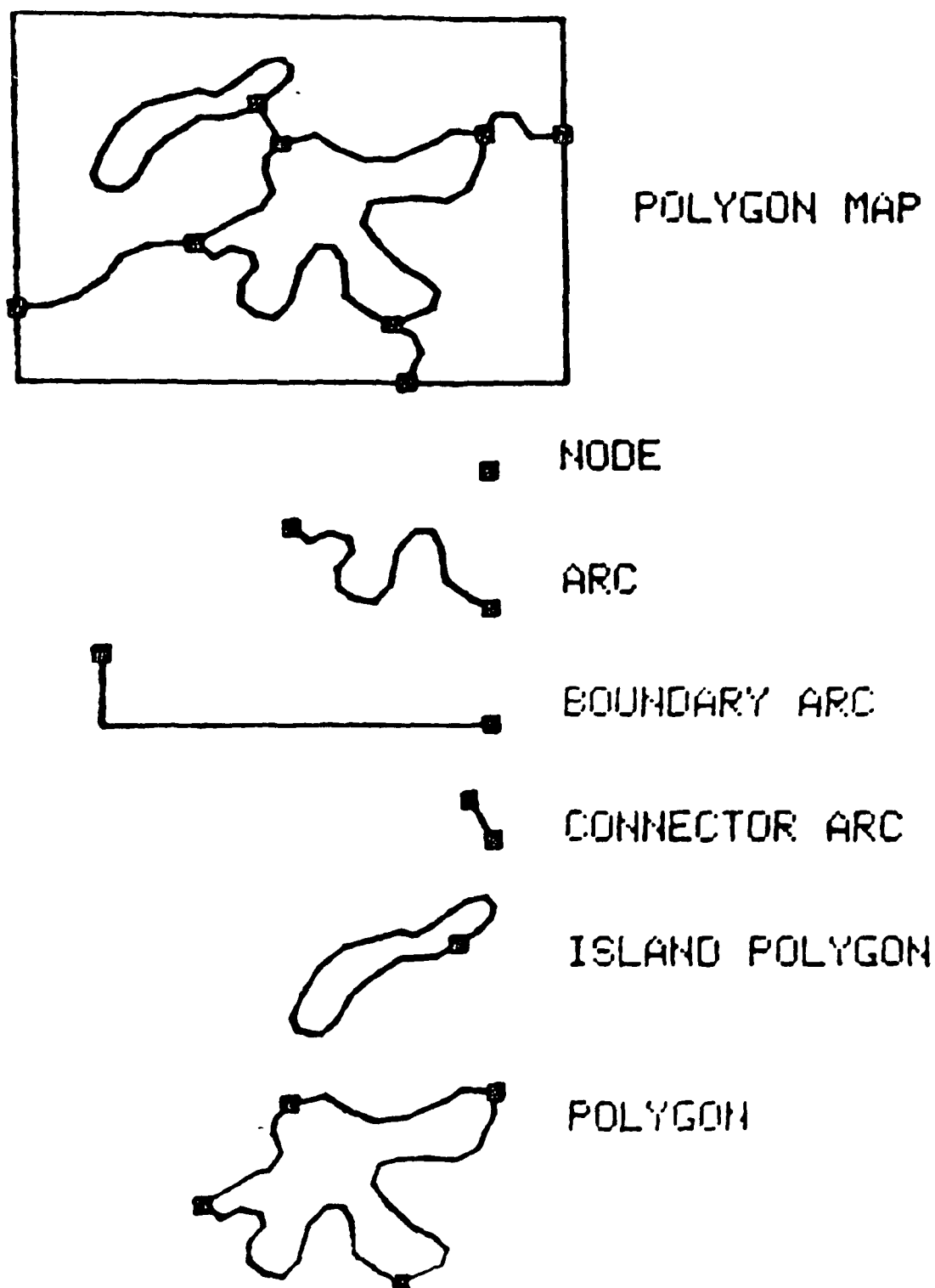


Figure 1. Topological elements of a polygon map.

Processing Digitizer Output

The X-Y coordinates of all mapped points are recorded using a Tektronix 4594 graphics tablet. This digitizer sends the coordinates of single points to the IBM 370/3033 every time a cursor pen is pressed down at the desired location on the soil map. Arcs and the points inside the polygons are digitized during different sessions. The end of an arc is entered by recording an X-Y value within a specific range recognized by the PROCESSOR program (see flow diagram in Fig. 2). If an arc or point is digitized incorrectly, an X-Y value is recorded within another specific range recognized by the software as an error flag. An initial processing program (PROCESSOR) reads in the digitized points, makes appropriate arc separations and arc or point deletions, and then creates portions of appropriately formatted arc or point files.

Table 1 shows the information contained in a complete arc file. All data except left and right polygon values are generated from digitizer output by the PROCESSOR program. Left and right polygon information is generated by the POLYGON program. Table 2 shows the information contained in a complete point-in-polygon file. Point flags, X and Y coordinates, and point numbers are generated by the PROCESSOR program. Polygon attributes and commentary information must be manually entered into the file.

Along with the number of X-Y coordinates, the number of topological elements may vary greatly between digitized maps. Factors which influence the number of coordinates include: the scale and area of the source map, the number and complexity of the mapping units, and the skill of the digitizer operator at choosing the minimum number of points to accurately represent the map. The number of mapping units is the major influence on the number of topological elements in a digital map. The number of islands and the complexity of the mapping units are additional factors.

Arc Cleaning

Files of arcs and points inside each mapping unit almost always contain errors when first digitized. Typical errors include: missing arcs, duplicate arcs, and two separate arcs recorded as a single arc passing through a node (Fig. 3). Because the resolution of the graphics tablet is 0.01 inches and the pen is hand-held, it is unlikely that arcs will end at precisely the same point. There is typically a 0.01 to 0.06 inch discrepancy between endpoints which should be identical. These errors are corrected by the CLEAN program without regard to the mapping unit represented by the individual arcs. The file of digitized points

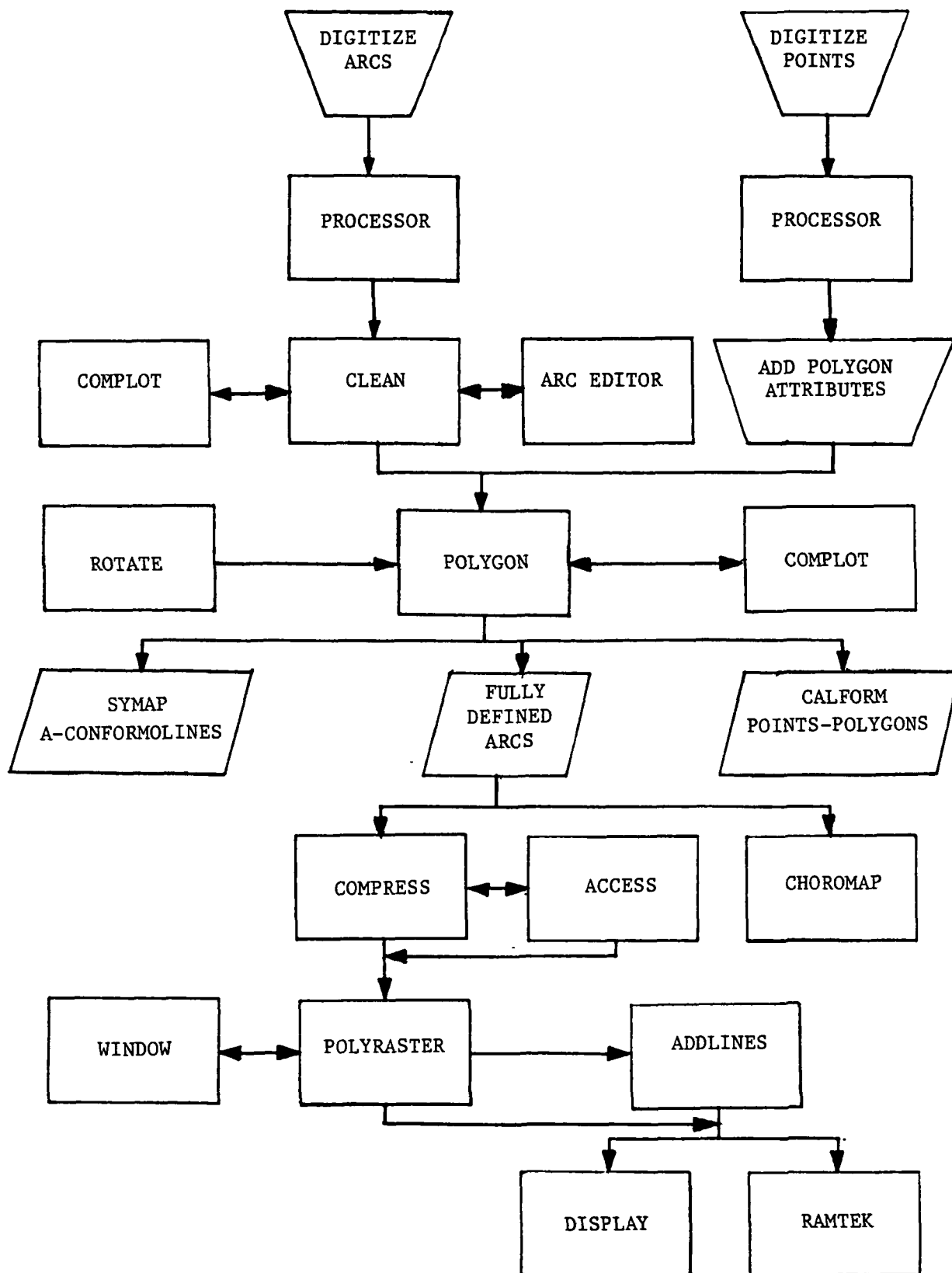


Figure 2. Flow chart of the geographic information system.

Table 1. Examples of Arc File Information

Flag indicating position on arc ^a	X Coordinate	Y Coordinate	Arc Number	Left Polygon	Right Polygon
3	830	1628	53	27	72
4	828	1632	53	27	72
5	823	1635	53	27	72
3	823	1635	54	56	72
4	811	1638	54	56	72
4	791	1637	54	56	72
5	753	1624	54	56	72
3	753	1624	55	72	55
4	752	1617	55	72	55
4	748	1611	55	72	55
4	739	1607	55	72	55
4	715	1603	55	72	55
4	645	1575	55	72	55
5	626	1565	55	72	55

^a3 = beginning, 4 = middle, 5 = end of arc

Table 2. Examples of Point File Information

Flag indicating the point is in a polygon	X Coordinate	Y Coordinate	Point Number	Attribute Code	Comments ^a
9	1461	1096	93	17	BMF
9	1484	1098	94	14	BkB
9	1496	1064	95	41	EeD
9	1512	1053	96	40	EgC
9	1448	1044	97	22	BrB
9	1460	1123	98	27	BxD
9	1510	1135	99	6	LoB
9	1532	1074	100	83	Ne
9	1551	1058	101	42	EgF
9	1577	1108	102	40	EgC

^aFor this table the commentary information is given in soil symbols.

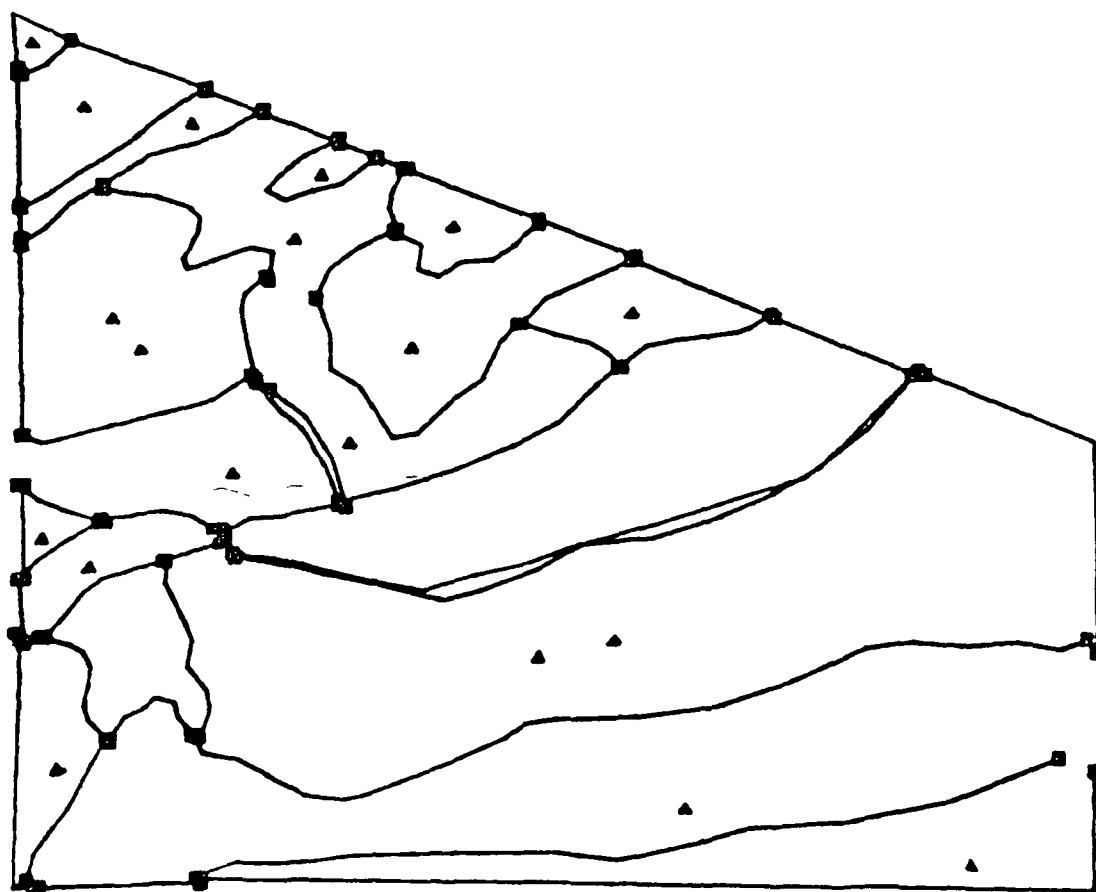


Figure 3. Plot of uncleaned arc and point data. Evident in this plot are missing arcs, duplicate arcs, missing points inside polygons, duplicate points inside polygons, and improperly linked arc endpoints.

inside each mapping unit can also contain initial errors. These include: duplicate points, missing points, and the wrong mapping unit appended to a particular point. Except for incorrectly labeled points, such errors are easily identified and corrected with the aid of the CLEAN program. Careful initial digitizing eliminates much of the need for error correction.

The CLEAN program can also be used to trap errors in the arc files. These errors include: nodes with too many or too few arc endpoints, duplicate arcs, and arcs which are too short. The program can also produce a file of the beginning and ending X-Y coordinates and the sequential node number of these points, as well as an updated coordinate file of the cleaned arcs.

The number of arc endpoints at a node is determined by searching for all endpoints which fall within a user-specified distance from each endpoint. If the distance is too large, nodes will be flagged with five, six, or more endpoints. An initial search radius of 0.034 inches has proved to be reasonable for the 0.01 inch resolution of the digitizer. A tabular listing of X-Y coordinates, arc number, and sequential number of the record in the file is produced for the nodes with too few or too many endpoints. This listing can be altered by the user. Missing arcs manifest themselves through nodes in the same general neighborhood which have too few arc endpoints. Obtaining a plot of the area in question is the best way to verify these clues. A summary of the number of cleaned nodes, node errors, and the number of nodes with two, three, four and more arc endpoints is also produced by the CLEAN program.

As the X-Y coordinates are read into the CLEAN program, the length of each arc is computed along with an average X and Y value for each arc. Arcs with length under a user-specified value (usually equal to the search radius used to check the number of arc endpoints) are flagged for possible deletion. Arcs are assumed to be duplicates if both have endpoints within a user-specified distance and the average of all coordinates in each arc is also within a user-specified distance. Figure 4 shows the same data as Fig. 3, after correction of arc and point-in-polygon errors.

Plotting

The COMLOT program was developed to generate plots of digitized arcs and/or points. This program may be used for display purposes (Fig. 5) or as an aid in trapping digitizing errors, such as those shown in Fig. 3. Plots are one of the best ways to verify the spatial accuracy of a digitizing effort. Plots may be generated for rapid display on CRT's of the Tektronix 4010 type, or for Tektronix 4662 flatbed plotters. Plot dimensions are approximately 6 by 8 inches and 10 by 15 inches, for the CRT and plotter, respectively.

Input to the COMLOT program consists of digital arc and/or point files and various options cards. Up to 20 separate plots may be generated in one run. The area covered in a plot is either the entire data set or



Figure 4. Plot of cleaned arc and point data for the same area shown in Fig. 3.



Figure 5. Plot of Atlas Sheet 8 from the Huntington County soil survey. (Data from Merkel, 1978)

a rectangular window within the data. Windows may be explicitly defined, or they may be computed from a user-specified center point and zoom distance. Arc number labels may be plotted on the basis of their sequential position in the input stream or their file labels. Point number labels may be plotted from their attribute code or from their position in the input stream. Boxes may be plotted at arc endpoints. A frame is usually also plotted, but this may be suppressed. It is possible to plot only user-specified arcs. If left and right polygon information is contained in the arc data, connector arcs may be suppressed and/or only user-specified polygons or only boundary arcs may be plotted.

Interactive Arc File Editor

To aid in the manual correction of arc file errors, an interactive file editor (ARC EDITOR) was developed on the IBM 370/3033, using the INTERACT pre-processor. With this editor, arcs can easily be merged, deleted, or split. It is also possible to list the file header; to list arcs, arc endpoints, or all coordinates with a particular attribute; and to resequence arc and coordinate numbers. Commands are given in the form of the desired operation, followed by the arc numbers to which it should be applied. The editor then issues sequences of commands to the computer to accomplish the desired operation. Possible errors are checked before actually performing any operation, so the editor is not only easier to use, but more reliable than entering individual commands from the keyboard. Many INTERACT commands, along with editor commands, may be entered while executing the editor, allowing a great degree of flexibility.

Polygon Formation

Once all of the arc errors are corrected and points have been recorded for each mapping unit, arcs are chained together into polygons. A polygon chaining program (POLYGON) was developed to link the appropriate arcs that bound each polygon. Determination of which polygon falls on the left and right hand side of each arc is made using information obtained from the point file. Polygons with multiple or missing points are trapped through topological error checking. Program options include the creation of files for use with various graphics software packages and display devices.

The POLYGON program takes a point from the point file, finds the closest arc, and then follows it and subsequent arcs in either a clockwise or counter-clockwise direction until the starting point is reached. The area of each polygon is calculated during this process. Any degree of polygon nesting is easily handled. Chaining of a polygon is aborted if a user-specified maximum number of arcs per polygon is exceeded or if a node is encountered with fewer than three arc endpoints. A limit on the number of arcs per polygon is set to avoid the endless chaining attempts which may result if arcs erroneously intersect near a node. The check on the number of arcs per node may be overridden, if desired. After

all points inside the polygons have been used, map perimeter arcs are chained and the area defined by the map perimeter is calculated. This is done for topological error checking.

Topological errors in the chaining process results from errors in digitizing or in coding attributes in the point files. These errors include multiple or missing points in a mapping unit polygon (see examples in Fig. 3) or the intersection of digitized lines. Since the perimeter is chained, each arc should define two, and only two, polygons; exceptions noted by the POLYGON program indicate errors. Errors are also trapped if the area defined by the map perimeter is not equal to the sum of the areas of the individual polygons. Duplicate polygons (those comprised of the same arcs) are flagged and missing polygons are indicated by arcs that define fewer than two polygons. If arcs intersect near a node, the wrong arc may be chained; this condition is also indicated by arcs that define the wrong number of polygons. The program can discern between connector arcs and arcs which erroneously have the same polygon attribute on both the left and right hand sides.

Tabular output from the POLYGON program includes a listing of the file statistics, the X-Y coordinates of each point inside a polygon, the associated polygon attributes, the arcs that compose each polygon, and the area of each polygon. If errors are found while chaining, the previous information is given for use in error correction, along with the X-Y coordinates where the chaining error occurred.

The following options for display and analysis are possible with the POLYGON program:

1. The left and right polygon numbers can be added to the arcs.
2. A file can be generated to show the arc numbers along with the left and right polygon numbers and the polygon attributes for each arc.
3. A file can be set up to contain the polygon numbers, attributes, and areas, and the component arcs of each polygon.
4. The left and right polygon attributes for each arc can be appended to the arc file for use with the Ramtek color display system.
5. An A-CONFORMOLINES package can be generated to permit the production of line-printer maps using the SYMAP program (Dougenik and Sheehan, 1975).
6. POINTS and POLYGONS packages can be generated for use with the CALFORM programs (Latham and White, 1978) or the CHOROMAP program.

Map Registration

Registration of a digitized map to a given base is frequently desirable. To help in registration, the ROTATE program was developed to perform transformations on digitized coordinates. Rotation can correct for improper alignment of a source map on a digitizer, for instance. Simple rotations can be done by inputting the degrees of rotation and, optionally, a pivot point. Coordinates may also be scaled to fit within a user-specified rectangular window. In this case, the spatial relationships between the X and Y coordinates are preserved, so that they will be tangent to a minimum of three sides of the window.

A rubber-sheet stretch can be performed by providing up to a fourth-order transformation polynomial to the ROTATE program. This polynomial can be obtained from the EZLS program (Myers, 1980), which performs a least-squares fit between sample points taken from a digital map and a given base. Output from EZLS consists of statistically significant parameters and detailed summary statistics regarding the goodness-of-fit of the model. Care must be taken when using this program to use a sufficient number of sample points and have a good point distribution, or the resulting polynomial could be useless for a transformation even though the statistics indicate a good model.

Compression and Decompression of Coordinates

It is desirable to use as little storage space as possible for digitized maps. Once digitized arcs have been cleaned and have left and right polygon values appended, they can be written into highly compressed files. Compression is based on two factors: removal of redundant information and binary formatting. The COMPLIT and POLYRASTER programs can accept input coordinates in compressed form.

Arc files have much redundant information that can be removed (see Table 1). This redundant information is included in order that the files can be more easily identified and edited by the user. The left and right polygon values, for instance, are the same for all points in an arc. Tags indicating the start or end of an arc are unnecessary if the number of X-Y coordinates in an arc are known. The sequential position of a particular arc in a file can be used instead of arc numbers. The minimal data necessary to describe an arc are: left and right polygon values, number of X-Y coordinates, and the actual X-Y values. Storage requirements for arc files can be further reduced if the data are written in binary format.

The COMPRESS and ACCESS programs were developed to compress or decompress digitized maps at the user's discretion. Compression can be done using either full or half words to represent X-Y coordinates. The left and right polygon attributes and the number of coordinates have values which can be stored in half-word integers on an IBM 370, so the total storage required for these three values is six bytes. Each X or Y coordinate normally requires four bytes of storage, but this requirement can be halved in some circumstances. The coordinate values used in arc

files are in digitizer units (or hundredths-of-inches, if a Tektronix 4594 is used). Fractional portions of these numbers, which would be beyond digitizer resolution, are not used. If the range in values of an arcs file in digitizer units is less than 65,536, half-word integers may be used to store each X-Y coordinate. The number of bytes required for storage of the coordinates in an arc is twice the number of coordinates times the number of bytes used in representing each coordinate.

Map File Subsets

It is sometimes desirable to work with only a portion of a digitized map. The WINDOW program was developed to produce a subset of a digital map file which contains left and right polygon information. Arcs or parts of arcs which fall inside the subset area are written to a new file. Boundary arcs with the appropriate left and right values are then generated for the perimeter of the subset area.

Polygon-to-Raster Conversion

The generation of thematic maps using certain devices, such as the Ramtek color display system, Optronics and Dicomed film recorders, or line printers, requires that polygon maps be converted into raster format.

Two types of output files may currently be produced from the POLYRASTER program, developed for polygon-to-raster conversion: Ramtek image files and the more versatile ORSER compressed raster maps. Each polygon attribute must be translated into one of 16 possible Ramtek codes or into one of 256 compressed map symbols. The DISPLAY program (Turner et al., 1978) can be used to generate graphics instructions for line printers, Optronics film recorders, Dicomed image recorders, or the Ramtek display system. Color interpretive maps may be generated for the Ramtek display system much more rapidly from compressed maps than from original polygon data, because only one scan of the polygon map is needed--that which generates the raster map file.

Input to the POLYRASTER program consists of a polygon map file, a translation table for polygon attributes, and control cards. Although the program accepts multiple input map files in compressed or uncompressed format, it is also possible to generate a raster file of one or more subsets of the input map(s). Because scratch disk space is used when necessary, very large polygon map files may be processed. This program can also justify text, and display titles and legend boxes included in the raster output file. The scale and pixel size of the raster image may be controlled by the user.

Raster conversion begins by reading the desired polygon map(s) into storage. Only arcs which have different left and right polygon values after translation and fall within or directly to the left of the desired map window are stored. If the polygons are not uniquely named, arcs to the left of the window are needed to ensure proper labeling in the raster map. Scan lines are generated at equal intervals, depending on

output map size. Intersections between a scan line and digitized arcs are stored for each of the scan lines. Scan lines are generated at intervals more precise than the resolution of the digitized arcs, to ensure that a scan line never intersects at a node. This scan line offsetting is similar to a method described by Broome (1977). Intersections are then sorted and the number and symbol for the pixels in each polygon are determined for the scan line. Actual raster output is generated by subroutines which write pixel count followed by pixel symbol in either Ramtek or ORSER compressed map format.

Color Display

Final color-interpretive maps can be produced from compressed map files using the Ramtek color display system. The RAMTEK program extended the existing graphic quality by the addition of justified text, legends, and reference color boxes to images. The ability to zoom in on a point was also implemented. Color symbols are assigned to the various mapping unit polygons by means of a translate table. Each mapping unit is given one of 16 possible color symbols based on a class attribute.

Output consists of a disk file of color graphics commands and a table showing the number of pixels, the area, and the percent of each symbol used in the image. Images are transmitted via a 1200 baud telephone line, from the disk file on the IBM 370/3033 to the Ramtek display system. Each of the 16 color symbols has an associated default color, but these may be interactively changed once the image is displayed on the color CRT screen. Each symbol may be assigned any of 4096 possible colors. Individual color schemes can be saved in a disk file for use at a later time. A hard copy of the interpretive maps can be obtained by photographing the screen.

Overlay of Ancillary Cartographic Data

The ADDLINES program was developed to overlay digitized ancillary line or point data onto a compressed raster map file. The program can be used for spatial referencing by overlaying, for example, a road network or state plane coordinate grid onto interpretive maps. It has also been used to overlay digitized soil boundaries onto a classified Landsat image (Imhoff et al., N.D.). The first step in the overlay process is to read the digitized ancillary lines and points, and their associated symbols, into computer memory. Scan lines from the raster map file are compared one-by-one with the digitized data. In the case of points, the closest pixel to the digitized point is changed to the symbol associated with that point. In the case of lines, changes are made to the symbols associated with the lines as follows: if the digitized line is perpendicular to the raster scan line, the intersection pixel is changed; if the digitized line intersects the raster scan line at less than 90 degrees, all scan line pixels less than one pixel-width from the digitized line are changed; if a digitized line is parallel to the raster scan line and within one pixel width of it, scan line pixels between endpoints of the digitized line are changed. After all changes have been made, raster scan lines are written into a new output file.

Interpretive maps can also be generated using devices other than the Ramtek. Line printer maps can be produced with the ORSER display program from raster map files, or with the SYMAP program (Dougenik and Sheehan, 1975) from files generated by the POLYGON program. Gray-scale plotter maps may be produced using the CALFORM program (Latham and White, 1978), from files generated by the POLYGON program. The DISPLAY program (Turner et al., 1978) may be used to generate image files from raster map files for use with Dicomed image recorders and Optronics film recorders.

Gray-Scale Display

The CALFORM program (Latham and White, 1978) yields high-quality gray-scale maps on Tektronix 4010 type terminals, Tektronix 4662 flatbed plotters, or CalComp 564 drum plotters. Because the version of CALFORM that is available at Penn State has been seriously degraded through source code modifications, the CHOROMAP program was developed to produce gray-scale interpretive maps on Tektronix 4010 CRT's or Tektronix flatbed plotters. In this program, the limit on the number of points in a polygon was set at 1000, which is adequate for most purposes. Shading capabilities consist of either single or cross-hatch lines which may be plotted at any angle. Legend capabilities are limited, but include automatic placement of map titles, key titles, and key legend box descriptions. Up to ten different categories may be used with either default or user-specified shading patterns. Coordinates are read into the program in standard uncompressed format and a table is provided to link each mapping unit to a shading category.

Determination of Area Statistics

The MAPCOMP program (Turner et al., 1978) was originally designed to allow two classification maps to be overlaid in order to map defined intersections and to determine area statistics. Thus, it can also be used to determine areas by classes within a polygon, if the polygon is stored in compressed raster map form, as output from the POLYRASTER program.

This process is demonstrated in Figures 6, 7, and 8. Figure 6 shows density-sliced Landsat data, generated using the NMAP program (Turner et al., 1978). A similar classification map might represent categories of forest defoliation. In Fig. 7, two polygons are shown, derived by digitizing the arcs comprising them and then processing the data through the series of programs described above. Such polygons could represent county or forest district boundaries. Since these outputs are both in ORSER compressed map format, they can be overlaid using the MAPCOMP program. The resulting output is shown in Fig. 8. Note that the areas outside the two polygons can be discarded. Tabular output from the MAPCOMP program provides a pixel count of all intersections of the two maps. These figures can easily be converted to area units for the classes enclosed by the polygons, as shown in Table 3.



Figure 6. Density-sliced Landsat scene, showing four categories. Such a map could represent four different degrees of forest defoliation, each assigned a unique color.

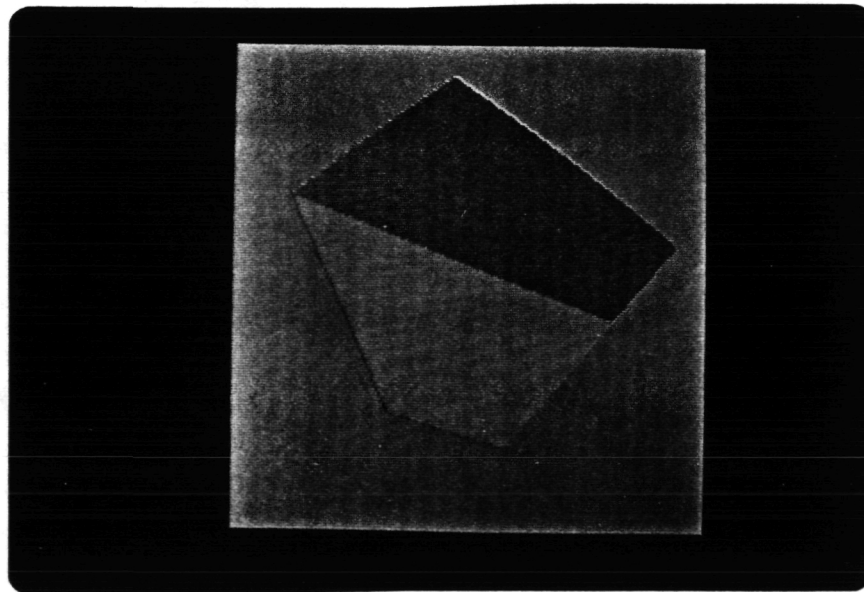


Figure 7. Two polygons derived by digitizing the boundaries and processing the resulting data through the PROCESSOR, CLEAN, POLYGON, COMPRESS, POLYRASTER, and DISPLAY programs. Such polygons could represent county or forest district boundaries.

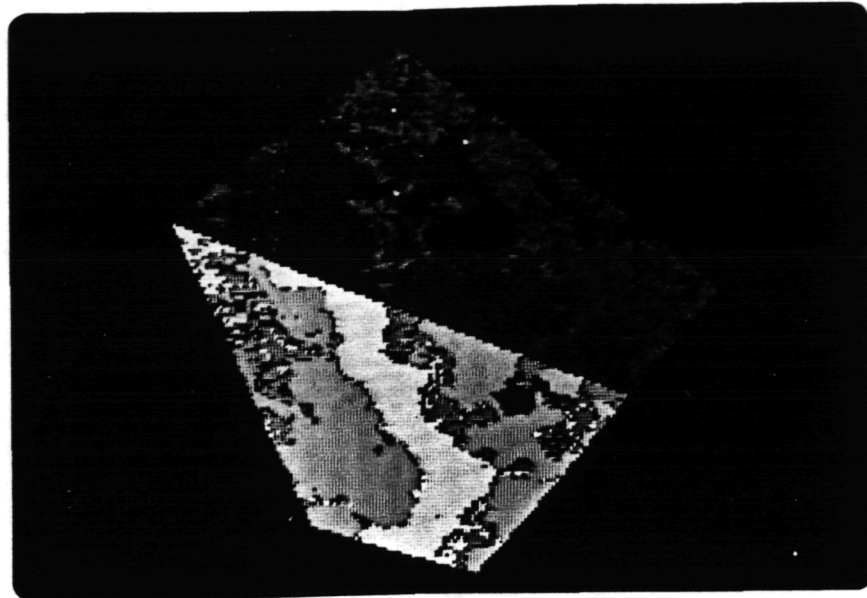


Figure 8. Overlay of the data shown in Figs. 6 and 7, using the MAPCOMP and DISPLAY programs.

Table 3. Tabulation of Acreages of Classes Within the Polygons Shown in Fig. 8 (Colors refer to that figure.)

<u>Polygon</u>	<u>Category</u>	<u>Color</u>	<u>Acres</u>
Top	A	Dark orange	2627
	B	Medium orange	2615
	C	Green	2155
	D	Yellow	1111
	Total		8508
Bottom	A	Light blue	1793
	B	Medium blue	3016
	C	Violet	1192
	D	Dark blue	896
	Total		6897

Feasibility of a Statewide Data Base/Information System

The study of the feasibility of a statewide data base/information system has resulted in some ideas being developed concerning the nature of a generalized geographic information system (GIS) which incorporates Landsat data. To reduce confusion with existing systems and programs, the conceptual interface between existing image processing systems, such as the ORSER system, and existing geographic data analysis systems, such as SYMAP (Dougenik and Sheehan, 1975) and MAP (Tomlin and Berry, 1979), has been given the acronym ZONAL which stands for "ZONation Algorithms."

The essential feature of this approach is an "indexing scene" which relates pixels in raster-format data to geographic referencing units used in other spatial data systems. The geographic referencing units are viewed conceptually as polygonal zones bearing index numbers. The polygons constitute a set of zones with either regular or irregular outlines, of which the rectangular cell is one special case. The index numbers may take the form of serial numbers in an ordered list of zones (e.g., an alphabetical list of counties), or of X-Y coordinates of zone reference points (perhaps centroids) in selected euclidean or non-euclidean planar coordinate systems.

The nature of the indexing scene is the result produced by raster-scanning a map of geographic referencing units and assigning each pixel a value corresponding to the index number of the zone in which the center of the pixel is situated. This pixel value takes the form of a k-element vector, with each element being one byte of indexing information. Options allow for assembling index values from these k bytes in several ways. One way is to select a sequential subset consisting of m bytes from the k bytes and composite the m bytes into a single index value. Another way is to select two sequential subsets of bytes and composite them into a pair of planar coordinates. The ability to specify subsets of the indexing vector provides for multiple layers of zones in the same indexing scene. The zones within any one layer must be mutually exclusive, but different layers can have overlapping zones. It is anticipated that a user-friendly executive program will be developed to allow easy access to any selected set of geographic referencing units or layer.

Geometric correction procedures are used to register Landsat or other raster-format data to the indexing scene. (The indexing scene may be merged as an extra set of "channels" if it becomes important to reduce the number of files.) The full range of Landsat analysis techniques may then be exercised on the registered scene. Classified scenes are used as input to programs in the ZONAL system, which effect linkages to other geographic information systems. Important functions required for such linkage are:

- 1) generation of indexing scenes,
- 2) tabulation of area percentage by classification symbol with zone,
- 3) overlaying zones between indexing layers to create new zones of intersection, and

4) preparation of output files compatible with the particular GIS.

Depending on the nature of the geographic referencing units, the indexing scene may be produced either as a mathematical partition of a map-projected surface or by digitizing a map of irregular polygonal zones. Options allow flagging edge zones with a negative sign to serve as a "partial zone" indicator.

The ZONAL interface for cell-oriented systems will consist of two primary programs. The first will simulate the action of a scanner, by scanning the grid of cells and outputting a "channel" of pixels. Instead of reflectance values, however, each pixel simply contains the identifier of the cell in which the center of the pixel is located. These pixels will be written on a computer tape, a disk file, or a deck of punched cards. Options will be available to accommodate any chosen size of rectangular referencing cell as well as any rectangular pixel. The function of this program will be similar to that of the POLYRASTER program, and it may be written as a modification of that program.

Input to the second program will consist of the file of pixelized cells along with the results of a pixel-by-pixel classification of the scene produced by the Landsat analysis system. This program will match the two files on a pixel-by-pixel basis, accumulating over each cell to obtain the percentage of the pixels in the cell that belong to specified categories, as determined by the Landsat classification. The output file will contain a cell-by-cell summary, formatted according to the requirements of the host GIS. The function of this program will be similar to that of the MAPCOMP program and it may be written as a modification of that program.

Cross-correlation of the Landsat layer with other layers residing in the data base is then accomplished with the analysis and display facilities of the GIS. Thus, the information extracted from Landsat can be used for overlays, updating, change detection, etc., without further recourse to the Landsat analysis system. Furthermore, pixelization of cells is a one-time operation that need not be repeated for later runs on additional Landsat layers, as long as the cellular referencing units are not altered.

CONCLUSIONS

The research and development mandated under this contract have been completed. Approximately half the programs in the ORSER system can now process full Landsat scenes in one pass, including programs in addition to those specified in the contract. We expect that all programs will be expanded by the end of June.

Through the development of a system for digitizing polygons, editing them and converting them to raster form, and the use of the existing MAPCOMP program, the ORSER system can now tabulate classification statistics for any irregularly-shaped polygon within a scene.

The feasibility of developing a data base/information system to incorporate Landsat data and ancillary data for the entire state of Pennsylvania has been examined. This has resulted in the conceptualization of an interface between the ORSER system and general purpose geographic data analysis systems. It has been demonstrated that such a system is certainly feasible and no major hindrances to the development of such a system in the next phase of this project are foreseen.

REFERENCES

- Broome, F. R. 1977. File structures and algorithms for the U.S. Bureau of the Census' automated statistical mapping. Paper presented at the Advanced Study Symposium on Topological Data Structures for Geographic Information Systems, Oct. 16-21. Laboratory for Computer Graphics and Spatial Analysis, Graduate School of Design, Harvard University, Cambridge, MA. (PAPERS/Notebook, Vol. 4)
- Dougenik, J. A., and D. E. Sheehan. 1975. SYMAP Users Reference Manual, Edition 5. Laboratory for Computer Graphics and Spatial Analysis, Graduate School of Design, Harvard University, Cambridge, MA.
- Fegeas, R. G. 1977. The graphic input procedure--an operational line segment/polygon graphic to digital conversion. Paper presented at the Advanced Study Symposium on Topological Data Structures for Geographic Information Systems, Oct. 16-21. Laboratory for Computer Graphics and Spatial Analysis, Graduate School of Design, Harvard University, Cambridge, MA. (PAPERS/Notebook, Vol. 4)
- Imhoff, M. L., G. W. Petersen, and S. G. Sykes. N.D. Digital overlay of cartographic information on Landsat MSS data for Soil Surveys. Paper submitted for publication in Photogrammetric Engineering and Remote Sensing.
- Latham, C. A., and D. White. 1978. CALFORM Manual. Laboratory for Computer Graphics and Spatial Analysis, Graduate School of Design, Harvard University, Cambridge, MA.
- Merkel, E. J. 1978. Soil Survey of Huntingdon County, Pennsylvania. U.S. Department of Agriculture, Soil Conservation Service, Washington, D.C.
- Myers, W. L. 1980. User's Guide to EZLS Regression Program. Information Report 106/OR. Office for Remote Sensing of Earth Resources, Institute for Research on Land and Water Resources, The Pennsylvania State University, University Park, PA.
- Tomlin, C. D., and J. K. Berry. 1979. A mathematical structure for cartographic modeling in environmental analysis. Proceedings, American Congress on Surveying and Mapping, 39th Annual Meeting, Washington, D.C. pp. 269-284.
- Turner, B. J., D. N. Applegate, and B. F. Merembeck. 1978. Satellite and Aircraft Multispectral Scanner Digital Data User Manual. Technical Report 9-78. Office for Remote Sensing of Earth Resources, The Pennsylvania State University, University Park, PA.